

# Genomic analysis of *Chromohalobacter salexigens*: clues about its carbon metabolism and the nature of its halophilic properties

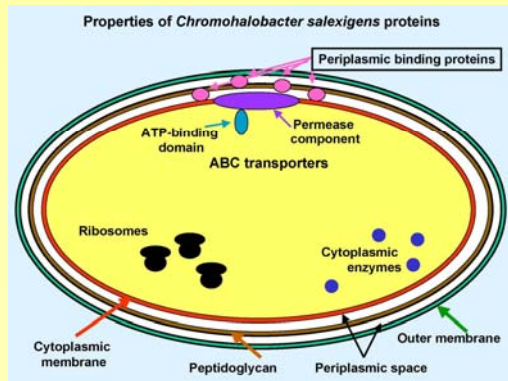
**Aharon Oren<sup>1</sup>, Laszlo N. Csonka<sup>2</sup>, Frank Larimer<sup>3</sup>, Paul Richardson<sup>4</sup>, Alla Lapidus<sup>4</sup>, Bradley W. Goodner<sup>5</sup> and Adam D. Ewing<sup>5</sup>**

<sup>1</sup>The Hebrew University of Jerusalem, Jerusalem, Israel; <sup>2</sup>Purdue University, West Lafayette IN, USA; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge TN, USA; <sup>4</sup>DOE Joint Genome Institute, Walnut Creek CA, USA; <sup>5</sup>Hiram College, Hiram, OH, USA

***Chromohalobacter salexigens* DSM 3043** is a moderately halophilic member of the  $\gamma$ -Proteobacteria. It grows at salt concentrations between 0.9 and 25% with an optimum at 7.5-10%. Pulsed-field gel electrophoresis of total DNA showed that *C. salexigens* possesses a chromosome of ~3.9 Mbp (63.9% G+C) and a low copy plasmid of < 100 kbp.

A draft sequence of the *C. salexigens* genome has been determined to an 8X coverage by the Joint Genome Institute of the US Department of Energy. Within the 3.7 Mbp unique sequences generated, 3370 predicted protein-coding genes were identified and provisionally annotated (<http://genome.jgi.psf.org/microbial/index.html>).

Over 60% of the ORFS show the highest similarity to orthologs from other  $\gamma$ -Proteobacteria. From the sequence, we find good indication that the organism has all the enzymes of glycolysis, hexose monophosphate shunt, Entner-Doudoroff pathway and the TCA cycle. We were able to rationalize the pathways of metabolism of many of the common sugars and mono- and dicarboxylic acids. We could account only partially for the metabolic pathway of benzoate, 4-hydroxybenzoate, and 3,4-hydroxybenzoate.



Compound	Enzymes in which it could be predicted from the sequence	Predicted enzymes in catabolic pathway that could not be found in the sequence	Can we account for all the secondary metabolic enzymes?
D-Glucose	several potential ABC and 10-coupled permeases, two glucokinases PQQ-dependent glucose oxidase glucosyltransferase K <sup>+</sup> coupled transporter glucokinase P <sub>1</sub> ATPase + P <sub>1</sub> TS Enzyme ABC <sup>+</sup>	hexokinase glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	Yes
D-Fructose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	galactose-1-P uridylyl transferase	No
D-Mannose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	Yes	
D-Galactose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	galactose-1-P uridylyl transferase	No
myo (1R,2S) inositol	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	2-6-5-epimerase kinase	No
Lactose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	β-galactosidase 6-P-β-galactosidase	No
Maltose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	amylomaltase	Maybe
Mannitol	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	periplasmic maltase	Maybe
Sorbitol	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	transport or phosphorylation system glucitol-1-P dehydrogenase	No
D-Glucoside (salivary)	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	glucosylate dehydrogenase 5-keto-4-D-glucuronate aldolase α-glucosidase (salivary)	No
D-Glucoside (saccharine)	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	arabinose + ribulose isomerase ribulose-5-P epimerase	No
L-Arabinose	glucose-6-phosphate kinase hexokinase phosphomannose isomerase hexokinase UDP-glucose hexokinase UDP-glucose-4-epimerase malic dehydrogenase, pyruvate malic 4-oxo-malate dehydrogenase pyruvate decarboxylase-4-oxo-malate fattylic acid	arabinose + ribulose isomerase ribulose-5-P epimerase	No

D-Ribose	ribokinase	xylose $\rightarrow$ xylobiose	Yes
D-Xylose	xylobiose kinase	xylobiose $\rightarrow$ xylobiose-5-phosphate	No
D-Erythritol	kinase?	erythritol $\rightarrow$ p-erythritol-5-phosphate erythritol $\rightarrow$ D-erythrose	No
TCA cycle intermediates	several potential "H <sup>+</sup> and H <sup>-</sup> linked to		
isocitrate	discharge and transporters, specificity can't be		
isocitrate	subsequent metabolism via TCA cycle		Yes
luminic acid			
Acetate	acetyl-CoA synthetase	acetyl-CoA synthetase	
	acetyl-CoA synthetase	acetyl-CoA synthetase	
	malate synthase	malate synthase	
	malate synthase	malate synthase	
	malate synthase (non-pyruvate specific)	malate synthase (non-pyruvate specific)	
Ethanol	alcohol dehydrogenase	alcohol dehydrogenase	Maybe
	alcohol dehydrogenase (non-pyruvate specific)	alcohol dehydrogenase (non-pyruvate specific)	
Glycerol	glycerol kinase	glycerol kinase	Yes
	glycerol-3-phosphate dehydrogenase	glycerol-3-phosphate dehydrogenase	Yes
D-Tartrate	tartrate dehydratase = $\alpha$ subunit	tartrate dehydratase = $\alpha$ subunit	Yes
D,L-Glycerate	D-glycerate kinase	D-glycerate kinase	Yes
	propionyl-CoA synthetase	propionyl-CoA synthetase	
	methionine synthase	methionine synthase	
Propionate	2-methylcrotonyl-CoA synthetase	2-methylcrotonyl-CoA synthetase	Yes
	malonyl-CoA ACP-ADP	malonyl-CoA ACP-ADP	
Malonate	malonate decarboxylase = $\beta$ subunit, poor match	malonate decarboxylase = $\beta$ subunit malonate decarboxylase- $\beta$ subunit malonate decarboxylase- $\beta$ subunit malonate decarboxylase- $\beta$ subunit	No

Compound	Enzymes in that could be predicted from the sequence	Predicted enzymes in catalytic pathway that could be found in the sequence	Can we account for all the necessary metabolic enzymes?
Berzote		See figure	Yes
Protocatechuate (3,4-dihydroxybenzoate)	protocatechuate 3,4-dioxygenase, 1; salutaridin 7-O-glucosyltransferase, 2; catabolite epimerase, 3; catabolite/methylcrotonate decarboxylase, 4	1. salutaridin 7-O-glucosyltransferase 2. catabolite epimerase 3. catabolite/methylcrotonate decarboxylase	Maybe not all can be metabolized
4-hydroxybenzoate	4-hydroxybenzoate 3-monoxygenase	1. salutaridin 7-O-glucosyltransferase 2. catabolite epimerase 3. catabolite/methylcrotonate decarboxylase 4. 4-hydroxybenzoate 3-monoxygenase	Yes
Toluene	1. salutaridin 7-O-glucosyltransferase 2. catabolite epimerase 3. catabolite/methylcrotonate decarboxylase 4. 4-hydroxybenzoate 3-monoxygenase 5. 2-hydroxy-4-isoxanate 2,4-diamine hydrolase 6. p-tyrosine 3-hydroxylase 7. p-tyrosine 3-hydroxylase 8. p-tyrosine 3-hydroxylase 9. p-tyrosine 3-hydroxylase 10. p-tyrosine 3-hydroxylase 11. p-tyrosine 3-hydroxylase 12. p-tyrosine 3-hydroxylase 13. p-tyrosine 3-hydroxylase 14. p-tyrosine 3-hydroxylase 15. p-tyrosine 3-hydroxylase 16. p-tyrosine 3-hydroxylase 17. p-tyrosine 3-hydroxylase 18. p-tyrosine 3-hydroxylase 19. p-tyrosine 3-hydroxylase 20. p-tyrosine 3-hydroxylase 21. p-tyrosine 3-hydroxylase 22. p-tyrosine 3-hydroxylase 23. p-tyrosine 3-hydroxylase 24. p-tyrosine 3-hydroxylase 25. p-tyrosine 3-hydroxylase 26. p-tyrosine 3-hydroxylase 27. p-tyrosine 3-hydroxylase 28. p-tyrosine 3-hydroxylase 29. p-tyrosine 3-hydroxylase 30. p-tyrosine 3-hydroxylase 31. p-tyrosine 3-hydroxylase 32. p-tyrosine 3-hydroxylase 33. p-tyrosine 3-hydroxylase 34. p-tyrosine 3-hydroxylase 35. p-tyrosine 3-hydroxylase 36. p-tyrosine 3-hydroxylase 37. p-tyrosine 3-hydroxylase 38. p-tyrosine 3-hydroxylase 39. p-tyrosine 3-hydroxylase 40. p-tyrosine 3-hydroxylase 41. p-tyrosine 3-hydroxylase 42. p-tyrosine 3-hydroxylase 43. p-tyrosine 3-hydroxylase 44. p-tyrosine 3-hydroxylase 45. p-tyrosine 3-hydroxylase 46. p-tyrosine 3-hydroxylase 47. p-tyrosine 3-hydroxylase 48. p-tyrosine 3-hydroxylase 49. p-tyrosine 3-hydroxylase 50. p-tyrosine 3-hydroxylase 51. p-tyrosine 3-hydroxylase 52. p-tyrosine 3-hydroxylase 53. p-tyrosine 3-hydroxylase 54. p-tyrosine 3-hydroxylase 55. p-tyrosine 3-hydroxylase 56. p-tyrosine 3-hydroxylase 57. p-tyrosine 3-hydroxylase 58. p-tyrosine 3-hydroxylase 59. p-tyrosine 3-hydroxylase 60. p-tyrosine 3-hydroxylase 61. p-tyrosine 3-hydroxylase 62. p-tyrosine 3-hydroxylase 63. p-tyrosine 3-hydroxylase 64. p-tyrosine 3-hydroxylase 65. p-tyrosine 3-hydroxylase 66. p-tyrosine 3-hydroxylase 67. p-tyrosine 3-hydroxylase 68. p-tyrosine 3-hydroxylase 69. p-tyrosine 3-hydroxylase 70. p-tyrosine 3-hydroxylase 71. p-tyrosine 3-hydroxylase 72. p-tyrosine 3-hydroxylase 73. p-tyrosine 3-hydroxylase 74. p-tyrosine 3-hydroxylase 75. p-tyrosine 3-hydroxylase 76. p-tyrosine 3-hydroxylase 77. p-tyrosine 3-hydroxylase 78. p-tyrosine 3-hydroxylase 79. p-tyrosine 3-hydroxylase 80. p-tyrosine 3-hydroxylase 81. p-tyrosine 3-hydroxylase 82. p-tyrosine 3-hydroxylase 83. p-tyrosine 3-hydroxylase 84. p-tyrosine 3-hydroxylase 85. p-tyrosine 3-hydroxylase 86. p-tyrosine 3-hydroxylase 87. p-tyrosine 3-hydroxylase 88. p-tyrosine 3-hydroxylase 89. p-tyrosine 3-hydroxylase 90. p-tyrosine 3-hydroxylase 91. p-tyrosine 3-hydroxylase 92. p-tyrosine 3-hydroxylase 93. p-tyrosine 3-hydroxylase 94. p-tyrosine 3-hydroxylase 95. p-tyrosine 3-hydroxylase 96. p-tyrosine 3-hydroxylase 97. p-tyrosine 3-hydroxylase 98. p-tyrosine 3-hydroxylase 99. p-tyrosine 3-hydroxylase 100. p-tyrosine 3-hydroxylase	No	

Comparison of the amino acid composition of different categories of proteins of *C. salexigens* and non-halophilic  $\gamma$ -Proteobacteria (*E. coli*, *P. aeruginosa*, *V. cholerae*) showed only a slight excess of acidic residues in the cytoplasmic proteins, and no significant differences were found in the acidity of membrane-bound proteins. In contrast, the periplasmic binding proteins of the ABC transport systems of *C. salexigens* have a pronouncedly lower mean pI value than the non-halophiles. *V. cholerae*, adapted to brackish water, shows intermediate values.

Gene category	<i>C. salexigens</i>	<i>E. coli</i>	<i>P. aeruginosa</i>	<i>V. cholerae</i>	<i>Halobacterium</i>
Central metabolism	5.10 ± 0.34 (12)	5.66 ± 0.49 (12)	5.80 ± 0.55 (12)	5.54 ± 0.44 (12)	4.22±0.14 (6)
Ribosomes	10.10 ± 2.15 (53)	10.41 ± 1.67 (53)	10.37 ± 1.81 (53)	10.28 ± 1.88 (53)	5.81 ± 2.56 (55)
ATP binding components	6.67 ± 1.47 (50)	6.93 ± 1.63 (73)	7.06 ± 1.62 (62)	6.90 ± 1.41 (54)	4.42 ± 1.10 (26)
Permease components	9.19 ± 1.56 (68)	9.18 ± 1.44 (77)	9.24 ± 1.49 (60)	8.55 ± 1.75 (53)	6.68 ± 2.46 (27)
Periplasmic binding components	4.54 ± 1.13 (55)	6.81 ± 1.56 (59)	7.28 ± 1.33 (44)	5.68 ± 1.03 (39)	4.11 ± 0.14 (6)

Ribosomes	Acidic	Basic	Acidic/Basic	Ser + Thr	Hydrophobic
C. satenigen	11.3	17.6	0.64	10.0	35.1
E. coli	10.9	10.9	0.80	9.6	28.1
P. aeruginosa	12.1	10.1	0.87	9.5	26.2
V. cholerae	10.7	12.9	0.85	9.5	27.8
Halobacterium	19.0	17.1	1.07	11.2	31.9
<b>Substrate binding proteins</b>					
C. satenigen	14.8	7.8	1.91	11.6	37.6
E. coli	15.1	10.2	1.58	10.1	37.4
P. aeruginosa	15.6	11.5	1.57	10.1	28.8
V. cholerae	15.4	9.7	1.17	12.3	37.8
<b>Enzymes of the central metabolism</b>					
C. satenigen	13.9	10.3	1.35	9.9	37.2
E. coli	12.8	10.7	1.19	10.8	37.5
P. aeruginosa	12.8	10.7	1.21	9.8	28.4
V. cholerae	12.4	10.4	1.20	10.2	38.2

'Halophilic' signatures of periplasmic binding proteins of ABC transporters, ribosomal proteins, and selected enzymes of the central metabolic pathways of *C. salexigens* as compared to *E. coli*, *P. aeruginosa*, *V. cholerae* and *Halobacterium* NRC-1. The values are given in mole-percent of the total number of amino acid residues or as ratios, as appropriate. Boldface: special features of the *C. salexigens* and the *Halobacterium* proteins.

Mean pI values of different categories of proteins of *Chromohalobacter salexigens*, as compared with the orthologs from the non-halophilic *Escherichia coli* K-12, *Pseudomonas aeruginosa* PA01, *Vibrio cholerae* O1 El Tor N16961, and the extremely halophilic archaeon *Halobacterium* sp. NRC-1.

## Conclusions

1. The sequence information shows that *C. salexigens* is a versatile heterotroph, and can at least partially metabolize a number of aromatic and xenobiotic compounds. This opens the possibility that the organism might be exploited for biological cleanup of highly saline polluted environments.
2. The acidic nature of the *C. salexigens* periplasmic substrate binding proteins is indicative of salt adaptation and possibly salt dependence of these proteins, and indicates that salt requirement of proteins located external to the cytoplasmic membrane may determine salt requirement of many prokaryotes.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, JINR Berkeley National Laboratory, University of Göttingen, DE, 46901-0519, and the Max-Planck-Gesellschaft, Institut für Experimentelle Kernphysik, Göttingen, DE, 34005, Germany.

LBNI-57689 Poster

## References

Csonka, L.N., O'Connor, K., Larimer, F., Richardson, P., Lapidus, A., Ewing, A.D., Goodner, B.W., and Oren, A. 2005. What we can deduce about metabolism in the moderate halophile *Chromohalobacter salexigens* from its genomic sequence. In: Gunde-Cimerman, N., Oren, A., and Plemenitis, A. (eds.). Adaptation to life at high salt concentrations in Archaea, Bacteria, and Eukarya. Springer. Dordrecht. in press.

Oren, A., Larimer, F., Richardson, P., Lapidus, A., and Csonka, L.N. 2005. How to be moderately halophilic with a broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. *Extremophiles*, in press

